

# AMPLICON LIBRARIES FOR ILLUMINA SEQUENCING

Updated October 31, 2017

Amplicon libraries are structured as follows:

```
5'-P5-index2-Rd1SP-NNNN-[your-primer-1]-[target]-[your-primer-2]-Rd2SP-index1-P7-3'  
3'-P5-index2-Rd1SP-NNNN-[your-primer-1]-[target]-[your-primer-2]-Rd2SP-index1-P7-5'
```

- P5/P7 = flow cell-binding sites
- Index2/1 = short, unique sequences used to identify each sample (aka i5/i7)
- Rd1SP/Rd2SP = read 1 and (if paired end) read 2 sequencing primer-binding sites
- NNNN = 4 random bases to increase base diversity across clusters (optional for projects using a small portion of a run or including multiple loci)

## SEQUENCING OPTIONS

The number of sequences required for your project will depend on:

- the number of samples
- the number of targets within each sample (e.g. multiplexed loci)
- the diversity/number of variant sequences within a sample
- the frequency of those variants

We routinely arrange shared MiSeq runs using the v2, 2x250 bp configuration. These runs will output approximately 30M paired end reads in total (15M templates sequenced from both directions). A small percentage (0.5-5%) of that output is sufficient for most samples. Please refer to our sequencing price list for other run configurations, but note that you will need to use the entire run.

## AMPLICON SIZE & ORIENTATION

The maximum size of your amplicon depends on your sequencing read length (e.g. 2x250 bp) and the positions of interesting SNPs, indels, etc. Your reads will begin immediately after the sequencing primer-binding sites, which means your target primers will be part of the output sequence, as will the optional NNNN bases.

For instance, 2x250 bp sequencing of a 400 bp target (plus two 20 bp target primers but no random bases) would look like:

```
Read 1                               5'====3'  
library      5'-P5-index2-Rd1SP====Rd2SP-index1-P7-3'  
              3'-P5-index2-Rd1SP====Rd2SP-index1-P7-5'  
Read 2                               3'====5'
```

Each "=" represents 10 bp

In this example the 3' ends of the two paired end reads will overlap by 60 bp in the middle of the target region. Since the base quality declines towards the 3' ends of the reads, this overlap may be necessary to produce a high quality consensus sequence. If the middle of your amplicon is not informative, you can increase your amplicon size and reduce the amount of overlap or even have a gap between reads.

In most cases it doesn't matter which of your target primers is attached to which adaptor. In fact, if you're having problems with non-specific product formation, try switching the orientation. If you are not performing paired end sequencing (read 1 only) or if amplicon length/sequence quality is problematic, be sure to orient the most critical part of your target nearest the P5 adaptor. You can expect higher quality data from read 1 than from read 2.

## PRIMER OPTION 1: target primers with complete adaptors

If you already have an amplicon like this:

```
5'-[your-primer-1]-[target]-[your-primer-2]-3'  
3'-[your-primer-1]-[target]-[your-primer-2]-5'
```

You can complete the library structure with one round of PCR (just a few cycles) using these primers:

```
5'-P5-[index2]-Rd1SP-NNNN-[your-primer-1]-3'  
5'-P7-[index1]-Rd2SP-[your-primer-2]-3'
```

Which are:

```
5'-AATGATACGGCGACCACCGAGATCTACAC-XXXXXXXX-ACACTCTTCCCTACACGACGCTCTCCGATCT-NNNN-[your-primer-1]-3'  
5'-CAAGCAGAAGACGGCATAACGAGAT-XXXXXXXX-GTGACTGGAGTTCAGACGTGTGCTCTCCGATCT-[your-primer-2]-3'
```

You can also use these primers directly on your gDNA/source material, though it might take some further optimizing and will require more cycles.

## PRIMER OPTION 2: target primers with partial adaptors and separate indexing primers

If you have multiple projects/pairs of target primers, it is more economical for your lab to purchase generic "indexing" primers that can be re-used with any target. This requires two rounds of PCR.

In the first round of PCR, only part of the adaptor sequence is added, which then serves as the binding site for the indexing primers in the second round. Only a few cycles are needed for the second round of PCR.

### 1st round primers:

```
5'-Rd1SP-NNNN-[your-primer-1]-3'  
5'-Rd2SP-[your-primer-2]-3'
```

Which are:

```
5'-ACACTCTTCCCTACACGACGCTCTCCGATCT-NNNN-[your-primer-1]-3'  
5'-GTGACTGGAGTTCAGACGTGTGCTCTCCGATCT-[your-primer-2]-3'
```

### 2nd round primers:

```
5'-P5-[index2]-Rd1SP-3'  
5'-P7-[index1]-Rd2SP-3'
```

Which are:

```
5'-AATGATACGGCGACCACCGAGATCTACAC-XXXXXXXX-ACACTCTTCCCTACACGACGCTCTCCGATCT-3'  
5'-CAAGCAGAAGACGGCATAACGAGAT-XXXXXXXX-GTGACTGGAGTTCAGACGTGTGCTCTCCGATCT-3'
```

## INDEXES

When setting up your PCRs, primers containing indexes are NOT added to your master mix. Each sample must have a unique combination of indexes, which means adding these primers separately to the reactions. For instance, a plate of 96 samples can be indexed using 20 primers: 8 variants of the i5 primer (one per row) and 12 variants of the i7 primer (one per column).

We recommend using Illumina's Nextera XT v2 index set, which offers 384 unique index combinations:

### i5 (index2):

Name	Sequence
S502	CTCTCTAT
S503	TATCCTCT
S505	GTAAGGAG
S506	ACTGCATA
S507	AAGGAGTA
S508	CTAAGCCT
S510	CGTCTAAT
S511	TCTCTCCG
S513	TCGACTAG
S515	TTCTAGCT
S516	CCTAGAGT
S517	GCGTAAGA
S518	CTATTAAG
S520	AAGGCTAT
S521	GAGCCTTA
S522	TTATGCGA

### I7 (index1):

Name	Sequence	Output
N701	TCGCCTTA	TAAGGCGA
N702	CTAGTACG	CGTACTAG
N703	TTCTGCCT	AGGCAGAA
N704	GCTCAGGA	TCCTGAGC
N705	AGGAGTCC	GGACTCCT
N706	CATGCCTA	TAGGCATG
N707	GTAGAGAG	CTCTCTAC
N710	CAGCCTCG	CGAGGCTG
N711	TGCCTCTT	AAGAGGCA
N712	TCCTCTAC	GTAGAGGA
N714	TCATGAGC	GCTCATGA
N715	CCTGAGAT	ATCTCAGG
N716	TAGCGAGT	ACTCGCTA
N718	GTAGCTCC	GGAGCTAC
N719	TACTACGC	GCGTAGTA
N720	AGGCTCCG	CGGAGCCT
N721	GCAGCGTA	TACGCTGC
N722	CTGCGCAT	ATGCGCAG
N723	GAGCGCTA	TAGCGCTC
N724	CGCTCAGT	ACTGAGCG
N726	GTCTTAGG	CCTAAGAC
N727	ACTGATCG	CGATCAGT
N728	TAGCTGCA	TGCAGCTA
N729	GACGTCGA	TCGACGTC

### Notes:

- insert indexes into your primers in the orientation shown in the "Sequence" column, replacing "XXXXXXXX" in the examples
- i5/index 2 must be between the P5- and Rd1SP-binding sites
- i7/index 1 must be between the P7- and R2SP-binding sites
- the "Output" of the the i7/index 1 read will be the reverse complement of the orientation in your primer; we will use the output orientation when de-multiplexing your samples and this will appear in your sequence headers. Include all columns with your sample submission information to avoid confusion

You may use any combinations of indexes, but consider:

- if you would like space on a shared MiSeq run, your index combinations must not match those used for other samples on a run. If there are index conflicts, your project will have to wait for the next available run. Before placing your primer order, ask us about currently popular indexes so you can avoid using those
- if you are not sharing a run, and have only a few samples that you are sequencing deeply, the index sequences need to be balanced. Please contact us for further instruction
- if you have only a small number of samples, you may only need to vary one of the primers. E.g. 4 samples could be indexed with 701, 702, 703, and 704, with 501 (in the PCR master mix) for all of them. Keep i5/index2 constant in these scenarios - i7/index1 should always be variable

### **ORDERING PRIMERS**

We recommend using IDT's "Ultramer" product for all primers containing indexes. If you prefer a different company, be sure to select additional purification to ensure full length products are obtained. The 1st round primers shown in Option 2 can be standard desalted oligos (length permitting).

### **PCR SET-UP**

Amplification conditions will depend on your choice of PCR reagents and your particular primers. You may need to optimize these conditions to minimize primer dimer formation. We recommend using a high fidelity polymerase such as NEB's Q5.

#### Tips:

- amplicon libraries typically require 20-35 total PCR cycles. If you will be checking your PCR products on a gel or using SequelPrep normalization, you will need sufficiently high DNA concentrations, but otherwise should aim to minimize the total number of PCR cycles
- it may be necessary to clean-up the reactions between the first and second rounds of PCR. This ensures that leftover round 1 primers do not compete with round 2 primers. Some users have been able to avoid extra clean-up by reducing round 1 primer concentrations and/or diluting the round 1 product going in to round 2
- when performing indexing PCR or otherwise re-amplifying an existing PCR product, remember that your input material is extremely concentrated. Ensure that this second round of PCR has an excess of primers to accommodate this much material and the number of additional cycles you will perform. If primers are not in excess, these final PCR cycles will result in templates denaturing and rehybridizing to other templates with the same priming sites, including primer dimers. If this happens, any bound dimers will not be removed during your clean-up steps and may not be obvious during our QC.
- to capture technical (PCR) variation, consider preparing multiple reactions for each sample then combining those prior to clean-up. It is not necessary to give these replicates unique indexes unless you need to measure technical variation. If biological replicates are needed in your analyses, these will need unique indexes
- multiple loci (i.e. pairs of target primers) can be multiplexed into a single reaction. If these combinations generate primer dimers or if different annealing conditions are required, try separating these into lower-plexity reactions. These can then be pooled together prior to indexing (i.e. all the loci for a particular sample)

## PURIFICATION AND POOLING

Your final library should be free of leftover primers, primer dimers, and other non-specific products, which will interfere with sequencing and QC. Even faint dimers may take up a large portion of your final dataset, as small products tend to be preferentially sequenced relative to larger products.

Given the large size of the Illumina adaptors, standard PCR clean-up procedures may not be effective. We recommend using AMPure XP Beads (or equivalent) which allow you to adjust your size selection cutoff. Gel excision may be necessary if your target product is small or similar in size to any non-target products. Please avoid heating the gel during your final extraction.

To simplify the procedure, you can pool your raw PCR products together and clean-up the final pool. If your product concentrations are variable and you want equal representation of all samples in your final dataset, you may need to normalize their concentrations prior to pooling. This can be done by quantifying the DNA (e.g. NanoDrop, fluorometer, estimating from gels, etc.) or using a product with limited binding capacity and removing excess material. We recommend SequalPrep Normalization Plates for this purpose and also offer this as a service.

Most DNA elution buffers are fine to use. Typically these are 10 mM Tris +/- 0.1 mM or 1 mM EDTA. We use 10 mM Tris with 0.1% Tween-20.

## SAMPLE SUBMISSION

Sample details must be provided by e-mail (to [fmf@mcmaster.ca](mailto:fmf@mcmaster.ca)) in an Excel or Open Office spreadsheet using the following columns:

- well locations or tube numbers/labels
- sample names (no spaces or special characters except hyphens and underscores)
- i5 indexes (ID and primer sequence)
- i7 indexes (ID, primer sequence, and output sequence)
- volume provided (ul)
- DNA concentration (ng/ul or nM) (not required for SequalPrep service)

Please include the following information in your e-mail and/or with your sample submission:

- your expected amplicon size, including the complete adaptor structure
- the read length/sequencing configuration you require (see our price list)
- for shared 2x250 bp runs: % of the run or number of reads per sample
- the BaseSpace account (e-mail address) to be used for data transfer
- Mosaic chartfields for billing (McMaster clients)

Tube, labelling, and concentration/volume requirements are different if SequalPrep normalization is required than if the libraries are ready for QC:

### For SequalPrep normalization

Amplicons must be submitted in 96-well PCR plates. Do not use flat bottom plates or PCR strip tubes. Partial plates must be filled by column (A01 to H01, then A02 to H02, etc.), not by row. Label both the edge of the plate and plate seal with your name, the date, and a project/plate name.

Successful normalization requires a minimum of 250 ng per well (e.g. 20 ul @ 12.5 ng/ul). We will draw equal volumes from each well using as much material as possible, up to 25 ul. Wells containing <250 ng may be underrepresented in the final pool.

### Library pools or individual libraries for QC

If you are submitting a pool or a small number of individual amplicons, use 1.5 ml tubes. Label the tops of each tube with your pool or sample names; keep your labels simple - you can list more detailed sample names in your spreadsheet. Include your name and the date on the side of the tube, as well as any project/batch name.

Amplicon concentrations should be no higher than 50 ng/ul. Low concentrations are acceptable. Please provide at least 20 ul.

### **DATA**

Your data will be uploaded to Illumina's BaseSpace site and ownership transferred to the account of your choice. If your data may not be uploaded to BaseSpace you must inform us in advance and make alternative arrangements.

The data will be de-multiplexed using the specified indexes and converted to FastQ format. For most runs you will receive two files for each sample: one containing all read 1 sequences and the other read 2 (if performed).

Full length sequences are provided. You will need to remove your target primer sequences and perform quality trimming as required for your analyses.

Please contact us for assistance with data analysis. Our bioinformatician can perform the analyses for you, develop a pipeline that you can use, or simply provide feedback/guidance.